

FIGURE 1A

1 CCCCCTGGAC AGCCGCCCTC TCCTCCAGGC CCGTGGGGCT GGCCCTGCAC CGCCGAGCTT CCGGGGATGA GGGCCCCCGG TGTGTCACC CGCGCGGCC
GGGGACCTG TCGCGGGAG AGGAGTCCG GGCACCCCGA CCGGACGTG GCGCTCGAA GGGCCCTACT CCGGGGGCC ACACCACTGG GCGCGCGGG

101 CAGGTGCTG AGGACCCCG GCGAGGGCG GAGATGGGG TGCACGGTGA GTACTCGCG GCTGGGGCT CCGCCCCCGG CGGTCCCTG TTTGAGCGGG
GTCCAGCGAC TCCCTGGGC CGTCCGGC CTCTACCCC ACGTGCCACT CATGAGCGC CGACCCGGA GGGCGGGCG GCCCAGGGAC AACTCGGCC

1 M G V H
^Exon 1
^Met

201 GATTAGCGC CCGGGCTATT GGCCGGGAGG TGGCTGGGT CAAGGACCG CGACTTGTC AGGACCCCG AAGGGGAGG GGGTGGGG AGCCTCCACG
CTAAATCGG GCGCCGATAA CCGGCCCTCC ACCGACCCAA GTTCTGGCC GCTGAACAGT TCCTGGGGC TTCCCTCTCC CCCCACCCG TCGGAGGTG

301 TGCCAGCGG GACTTGGGG AGTCTTGGG GATGGGAAA ACCTGACCTG TGAAGGGGAC ACAGTTTGG GGTGAGGGG AAGAAGGTTT GGGGTTCCTG
ACGTCGCC CTGAACCCC TCAGGAACCC CTACCGTTT TGGACTGGAC ACTTCCCTG TGCAAAACC CCAACTCCC TTCTTCCAA CCCCCAAGAC

401 CTGTGCCAST GGAGAGGAAG CTGATAAGCT GATAACCTGG GCGCTGGAG CACCACCTAT CTGCCAGAG GNNNTGGTA GCTGGGGTG GGTGTGCAC
GACACGGTSA CCTCTCTTC GACTATTGGA CTATTGGACC CCGGACCTG GTGGTGAATA GACGTCTCC CNNNNACCAT CGACCCCCAC CCCACACGTG

501 ACGGCAGCAG GATTGAATGA AGGCCAGGGA GGCAGCACCT GAGTGCTTG GAGTGGGG ACAGGAAGGA CGAGCTGGG CAGAGACGTG GGGATGAAGG
TGCCGTCGTC CTAACCTACT TCCGGTCCCT CCGTCGTGGA CTCACGAAC TACCAACCC TGTCCTCTCT GCTCGACCC GTCTCTGCAC CCTACTTCC

601 AAGCTGTCCT TCCACAGCCA CCCTTCTCCC TCCCGCCTG ACTCTCAGC TGGCTATCTC TTCTAGAATG TCCTGCCCTG CTGTGGCTC TCCTGTCCCT
TTCGACAGGA AGGTGTCGT GGAAGAGGG AGGGCGGAC TGAGAGTCGG ACCGATAGAG AAGATCTTAC AGGACGGACC GACACCGAAG AGGACAGGA

1 E C P A W L W L L S L
^exon 2

701 GGTGTGCTC CTCTGGGCC TCCAGTCTCT GGGGCCCCCA CCAGCCTCA TCTGTACAG CCGAGTCTG GAGAGTACC TCTTGAGGC CAAGAGGCC
CGACAGCGAG GGAGACCCG AGGTACAGGA CCGCGGGGT GGTGCGGAGT AGACACTGTC GGCTCAGGAC CTCTCCATG AGAACCTCG GTTCTCCCG

13 L S L P L G L P V L G A P P R L I C D S R V L E R Y L L E A K E A

801 GAGATATCA CGGTGAGACC CCTTCCCCAG CACATTCCAC AGAATCAG CTCAGGGCTT CAGGGAACCTC CTCCCAGATC CAGGAACCTG GCACCTGGTT
CTCTTATAGT GCCACTCTG GGAAGGGGTC GTGTAAGGTG TCTTGAGTGC GAGTCCCGAA GTCCCTTGAG GAGGTCTAG GTCCTTGAC CGTGAACCAA

46 E N I T

FIGURE 1B

901 TGGGGTGGAG TTGGGAAGCT AGACACTGCC CCCCTACATA AGAATAAGTC TGGTGGCCCC AAACCATACC TGGAACTAG GCAAGGAGCA AAGCCAGCAG
ACCCACCTC AACCCTTCCA TCTGTGACGG GGGGATGTAT TCTTATTGAG ACCACGGGG TTGGGTATGG ACCTTTGATC CGTTCCTCGT TTCGGTCTGC

1001 ATCCTACGGC CTGTGGGGCC A GGGCTGTC TGCAATTCAG ACGGGCTGTG CCGAACACTG CAGCTTGAAT
TAGGATGCCG GACACCCGGT CCCGGTCTCG GGAAGTCCCTG GGGCCGACAC ACCTAAAGTC TGCCCGACAC GCCTTGAGC GTCGAACTTA
T G C A E H C S L N
^exon 3

1101 GAGATATCA CTGTCCCAGA CACCAAAGTT AATTCTATG COTGAAGAG GATGGAGTG AGTTCCTTT TTTTTTTTT TCCTTCTTT TGAGAATCT

11 EN T T V P D T K V N F Y A W K R M E

CTCTTATAGT GACAGGTCT GTGGTTTCAA TTAAAGATAC GGACTTCTC CTACCTCCAC TCAAGGAAA AAAAAAAA AGGAAGAAA ACCTCTTAGA

1201 CATTTGGAG CCTGATTTTG GATGAAAGG AGAATGATCG AGGGAAGGT AAAATGGAGC AGCAGAGATG AGGCTGCCCTG GGCGCAGAGG CTCACGTCTA
GTAACGCTC GGACTAAAC CTACTTTCCC TCCTACTAGC TCCCTTTCCA TTTTACCTCG TCGTCTCTAC TCCGACGGAC CCGCGTCTCC GAGTGCAGAT

1301 TAATCCCAGG CTGAGATGGC CGAGATGGGA GAATTGCTTG AGCCTCGAG TTTCAGACCA ACCTGGGCAG CATAGTGAGA TCCCCCATCT CTACAAACAT
ATTAGGTCC GACTCTACCG GCTCTACCC CTTAACGAAC TCGGACCTC AAAGTCTGGT TGGACCCGTC GATCACTCT AGGGGGTAGA GATGTTTCTA

1401 TTAAAAAAT TAGTCAGGT AGGTGGTGCA TGGTGGTAGT CCCAGATATT TCGAAGGCTG AGGCGGGAGG ATCGCTTGAG CCCAGGAATT TGAGGTGCA
AATTTTTTA ATCAGTCCAC TCCACCACGT ACCACCATCA GGGTCTATAA ACCTTCCGAC TCCGCCCTCC TAGCGAACTC GGGTCTTAA ACTCCGACGT

1501 GTGAGCTGTG ATCACCAC TCACTCCAG CCTAGTGAC AGAGTGAGC CCTGTCTCA AAAAGAAAAG AATAATGAGG GCTGTATGGA
CACTCGACAC TAGTGTGGT ACGTGAGGTC GGAGTCACTG TCTCACTCCG GGACAGAGTT TTTTCTTTTC TTTTCTTTT TTATTACTCC CGACATACCT

1601 ATACATTCA TATTCATCA CTCACATCAT CATTCAATCA TTCATTCCNN NNNTCTTATT GCATACCTCT GTTTGCTCAG CTTGGTGCTT GGGGCTGCTG

1701 AGGGGCAGGA GGGAGAGGT GGCATGGGTC AGCTGACTCC CAGAGTCCAC TCCCTGTAGG TCAGGCAGCA GGCCGTAGAA GTCTGGCAGG GCCTGGCCCT
TCCCGTCCT CCCTCTCCCA CCGTACCCAG TCGACTGAGG GTCTCAGGTG AGGACATCC AGTCGGTCGT CCGGCATCTT CAGACCTCC CGGACCGGGA
V R Q Q A V E V W Q G L A L
^exon 4

1801 GCTCTCGGAA GCTGTCTCTGC GGGGCCAGGC CCTGTTGGTC AACTCTTCCC AGCGTGCGGA GCCCTGCAG CTGCATGTGG ATAAAGCCGT CAGTGGCCTT
CGAGAGCCTT CGACAGGACG CCCCCGTCCG GGACAACCAG TTGAGAAAGG TCGCACCCCT CGGGGACGTC GACGTACACC TATTTGGCA GTCACCGGAA

15 L S E A V L R G Q A L L V N S S Q P W E P L Q L H V D K A V S G L

1901 CGCAGCCTCA CCACTCTGCT TGGGCTCTG GGAGCCGAGG TGAGTAGGAG CGGACACTTC TGCTTGCCCT TTCTGTAAGA AAGGGAGAAG GGTCTTGCTA
CGCTCGGAGT GGTGAGACGA AGCCGAGAC CCTCGGGTCC ACTCATCCTC GCTCTGAAG ACGAACGGGA AAGACATTCT TTCCCTCTTC CCAGAACGAT

FIGURE 1C

2001 AGGAGTACAG GAACTGTCCG TATTCCTTCC CCTTCTGTGG CACTGCAGCG ACCTCCTGTT TTCTCCTTGG CAGAAGGAAG CCATCTCCCC TCAGATGCG
 TCCTCATGTC CTTGACAGGC ATAAGGAAGG GGAAGACACC GTGACGTCCG TGGAGGACAA AAGAGGAACC GTCTTCCTTC GGTAGAGGG AGGTCTACGC
 K E A I S P P D A
 ^exon 5

1

2101 GCCTCAGCTG CTCCACTCCG AACATCACT GCTGACACTT TCCGCAAACT CTTCGAGTC TACTCCAATT TCCTCCGGG AAAGCTGAAG CTGTACACAG
 CGGAGTCGAC GAGGTGAGGC TTGTTAGTGA CGACTGTGAA AGGCGTTTGA GAAGGTTAA ATGAGGTTAA AGGAGGCCCC TTTCGACTTC GACATGTGTC
 10 A S A A P L R T I T A D T F R K L F R V Y S N F L R G K L K L Y T G

2201 GGGAGGCCTG CAGGACAGG GACAGATGAC CAGGTGTGTC CACCTGGGCA TATCCACCAC CTCCTCACC AACATTGCTT GTGCCACACC CTCCTCCGCG
 CCCTCCGGAC GTCCTGTCCC CTGTCTACTG GTCCACACAG GTGGACCCCGT ATAGGTGGTG GAGGAGTGG TTGTAACGAA CACGGTGTGG GAGGGGGCGG
 44 E A C R T G D R O

2301 ACTCCTGAAC CCCGTCGAGG AGCTCTCAG
 TGAGGACTTG GGGCAGCTCC TCGAGAGTC

FIGURE 2

1 ATGGGGGTGC AGAATGTCC TGCCTGGCTG TGGCTTCTCC TGTCCTGTCT GTCCCTCCCT CTGGGCCCTCC CAGTCCTGGG CGCCCCACCA CGCCTCATCT
TACCCCCACG TGCTTACAGG ACGGACCGAC ACAGGAGAGG ACAGGAGCGA CAGCGAGGGA GACCCGGAGG GTCAGGACCC CGCGGTGGT GCGGAGTAGA
1 M G V H E C P A W L W L L L S L L S L P L G L P V L G A P P R L I C
101 GTGACAGCGG AGTCCTGGAG AGGTACCTCT TGGAGGCCAA GGAGGCCGAG AATATCAGCA CCGCTGTGC CGAACACTGC AGCTTGAATG AGAATATCAC
CACTGTCCGC TCAGGACCTC TCCATGGAGA ACCTCCGGTT CCTCCGGCTC TTATAGTCTT GCGGACACG CTTGTGACG TCGAACTTAC TCTTATAGTG
35 D S R V L E R Y L L E A K E A E N I T T G C A E H C S L N E N I T
201 TGTCCACAGC ACCAAGTTA ATTTCTATGC CTGGAAGAGG ATGGAGGTCA GGCAGCAGGC CGTAGAAGTC TGGCAGGGCC TGGCCCTGCT CTCGGAAGCT
ACAGGCTCTG TGGTTTCAAT TAAAGATACG GACCTTCTCC TACCTCCAGT CCGTCGTCCG GCATCTTCAG ACCGTCCCG ACCGGGACGA GAGCCTTCGA
68 V P D T K V N F Y A W K R M E V R Q Q A V E V W Q G L A L L S E A
301 GTCCTGCGGG GGCAGGCCCT GTTGGTCAAC TCTTCCCAGC CCGTGGAGCC CCTGCAGCTG CATGTGGATA AAGCCGTGAG TGGCCTTCCG AGCCTCACCA
CAGGACGCCC CCGTCCGGGA CAACCAAGTG AGAAGGTCG GCACCTCCG GCACGTCGAC GTACACCTAT TTCGGCAGTC ACCGGAAGCG TCGGAGTGGT
101 V L R G Q A L L V N S S Q P W E P L Q L H V D K A V S G L R S L T T
401 CTCTGCTTCG GGCTCTGGGA GCCCAGAAGG AAGCCATCTC CCCTCCAGAT CCGGCCCTCAG CTGCTCCACT CCGAACAAATC ACTGCTGACA CTTTCCGCAA
GAGACGAAGC CCGAGACCCCT CCGGTCTTCC TTCGGTAGAG GGGAGGTCTA CCGCCGAGTC GACGAGGTGA GGCTTGTAG TGACGACTGT GAAAGGCGTT
135 L L R A L G A Q K E A I S P P D A A S A A P L R T I T A D T F R K
^Polymorphism in CRL 1609 (A or C). A changes the protein sequence to K.
501 ACTCTTCCGA GTCTACTCCA ATTTCTCTCC GGGAAAGCTG AAGCTGTACA CAGGGGAGGC CTGCAGGACA GGGACAGAT GACCA
TGAGAAGGCT CAGATGAGGT TAAAGGAGGC CCCTTCGAC TTCGACATGT GTCCCTCCG GACGTCCTGT CCCCTGTCTA CTGGT
168 L F R V Y S N F L R G K L K L Y T G E A C R T G D R O

FIGURE 3

human MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE

chepo MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE

human NITTGCAEHCSLNENITVPDTKVNIFYAWKRMEVGQQAVEVWQGLALLSEA

chepo NITTGCAEHCSLNENITVPDTKVNIFYAWKRMEVRQQAVEVWQGLALLSEA

human VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLRLALGAQKEAISPPD

chepo VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLRLALGAXKEAISPPD

human AASAAPLRTITADTFRKLFRVYSNFLRGKCLKLYTGEACRTGDR

chepo AASAAPLRTITADTFRKLFRVYSNFLRGKCLKLYTGEACRTGDR

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 4

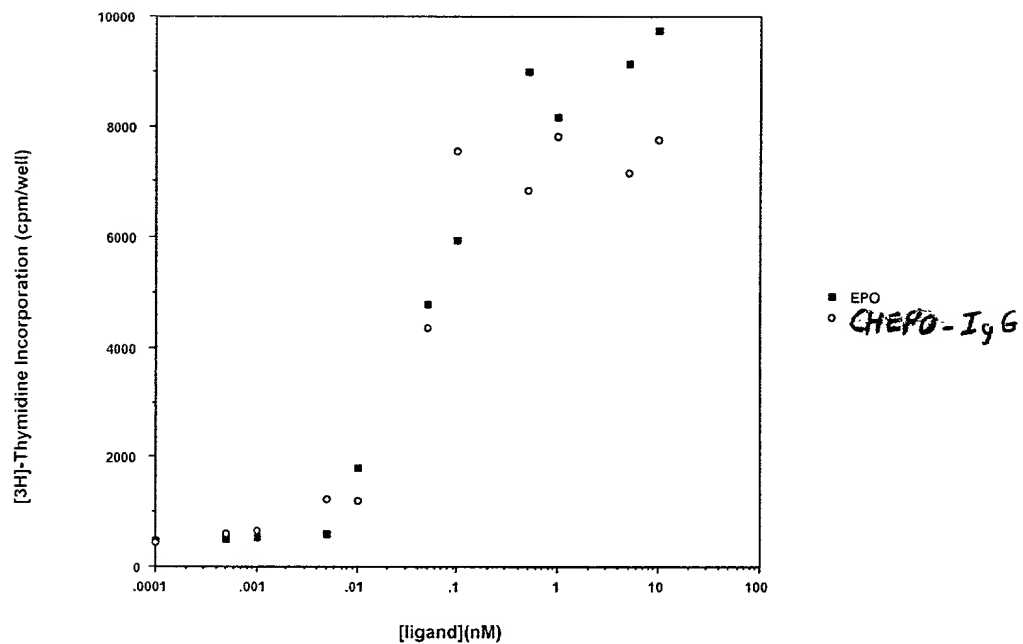


FIGURE 5

